



26

27

28

29

30 **Abstract:**

31 The Covid-19 pandemic has changed the paradigms for disease surveillance and rapid  
32 deployment of scientific-based evidence for understanding disease biology, susceptibility,  
33 and treatment. We have organized a large-scale genome-wide association study in Sars-  
34 Cov-2 infected individuals in Sao Paulo, Brazil, one of the most affected areas of the  
35 pandemic in the country, itself one of the most affected in the world. Here we present the  
36 results of the initial analysis in the first 5,233 participants of the BRACOVID study.

37 We have conducted a GWAS for Covid-19 hospitalization enrolling 3533 cases  
38 (hospitalized Covid-19 participants) and 1700 controls (non-hospitalized Covid-19  
39 participants). Models were adjusted by age, sex and the 4 first principal components. A  
40 meta-analysis was also conducted merging BRACOVID hospitalization data with the  
41 Human Genetic Initiative (HGI) Consortia results.

42 BRACOVID results validated most loci previously identified in the HGI meta-analysis.  
43 In addition, no significant heterogeneity according to ancestral group within the Brazilian  
44 population was observed for the two most important Covid-19 severity associated loci:  
45 3p21.31 and Chr21 near IFNAR2. Using only data provided by BRACOVID a new  
46 genome-wide significant locus was identified on Chr1 near the genes DSTYK and  
47 RBBP5. The associated haplotype has also been previously associated with a number of  
48 blood cell related traits and might play a role in modulating the immune response in  
49 Covid-19 cases.

50

51

52

53

## 54 **Introduction**

55

56 The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was first  
57 described in Wuhan, China, in December 2019, and quickly spread across the globe  
58 causing a pandemic (1). After 20 months of the beginning of the spread, there were more  
59 than 213 million confirmed cases worldwide, with total deaths exceeding 4 million, half  
60 million in Brazil alone (2).

61 Clinical outcomes of SARS-CoV-2 infection range from asymptomatic infection  
62 to fatal coronavirus disease 2019 (Covid-19). The high mortality rates are predominantly  
63 among specific subgroups with predeterminate conditions, such as male sex, older age or  
64 having pre-existent cardiovascular, pulmonary, or immune disease. However, many  
65 severe clinical cases are not due to comorbidity or constitutional factors. Notably, even  
66 among those without pre-existing conditions, inter-individual heterogeneity is striking.  
67 The considerable variation in disease behavior among infected individuals highlights the  
68 importance of the hosts' genetic background modulating disease presentation (3–5).

69 Until now a small number of large-scale genome-wide association studies  
70 (GWAS) has been conducted with some reproducible identified genetic loci associated  
71 with disease susceptibility or severity. A GWAS among 1980 patients from the European  
72 Covid-19 epicenters in Italy and Spain, as well as from the less-burdened countries of  
73 Germany and Norway was the first study to describe the susceptibility locus at the  
74 chromosome 3p21.31 gene cluster and suggested the involvement of the ABO blood-  
75 group system in Covid-19 (6). The *GenOMICC* study, in addition to replicating the  
76 association at chromosome 3p21.31 as mediator of critical illness, identified associations  
77 with disease severity and other genetic variants (7). Finally, early in 2021, the Covid-19  
78 Host Genetics Initiative (HGI) has brought together 49,562 cases and 2 million controls,  
79 from populations with different ancestries across the world to elucidate the genetic  
80 determinants of SARS-CoV-2 infection and severity including 853 samples from the  
81 BRACOVID study. From these analyses 15 genome-wide significant loci associated with  
82 SARS-CoV-2 infection or Covid-19 were observed (8).

83

84 Despite these efforts, the still limited number of identified loci that appears to  
85 modulate the relationship between clinical severity and human genetic variation  
86 precludes a thoroughly understanding of inter-individual susceptibility in the disease. In  
87 addition, most studies are overwhelmingly represented by European or Asian individuals  
88 and whether the identified loci are also operant on individuals from other ancestry  
89 backgrounds is unknown. Here we present the results from the BRACOVID study  
90 enrolling 5,233 participants with distinct COVID-19 clinical presentations and from a  
91 diverse population in Brazil. Our results confirm most previously identified loci and add  
92 a new genome-wide significant locus associated with COVID-19 hospitalization in the  
93 Brazilian population.

94

## 95 **Results**

96

97 The BRACOVID study was designed to identify and characterize genetic risk  
98 factors for Covid-19 severity in the Brazilian population.

99

### 100 **Relationship between demographic and clinical factors with Covid-19** 101 **hospitalization**

102

103 We summarize the demographic, clinical and laboratory characteristics of the  
104 5,233 participants of the study in Table 1 according to hospitalization necessity.

105

### 106 **Genome-Wide Association Analysis of Covid-19 hospitalization**

107

108 We have performed a GWAS of the genetic association architecture regarding  
109 Covid-19 hospitalization need. Groups were 1,700 individuals that did not require  
110 hospital admission due to Covid-19 and 3,533 individuals hospitalized due to severe  
111 Covid-19. In the primary analysis, we have run 2 GWAS analysis (one for each used  
112 genotyping array). Each analysis was adjusted for the first 4 Principal Components, age,  
113 and sex. GWAS results were then merged using the meta-analysis routine in plink. The  
114 observed genomic inflation factor for the adjusted analysis was  $\lambda = 1.05$ .

115 In our main analysis, we observed one genome-wide significant locus that reached  
116 the pre-defined genome-wide significant level of  $5 \times 10^{-8}$  (Figure 1). We also observed  
117 four additional loci that reached the pre-defined p-value threshold of  $1 \times 10^{-6}$  (Table 2  
118 and Supplementary Figure 1). We have run sensitivity analysis stratifying individuals by  
119 their main genetic ancestral component and then conducting a trans-ethnic meta-analysis.  
120 Results were very similar to those observed in the main analysis (Supplementary Figures  
121 2 and 3).

122

### 123 **BRACOVID HGI meta-analysis:**

124

125 We have replicated most previously described HGI (8) (B1 analysis) results  
126 (BRACOVID was not part of this specific HGI sub-analysis) (Supplementary File 2).  
127 Specifically, both loci at 3p21.31 and Chr21 near IFNAR2 previously associated with  
128 hospitalization need on HGI B1 analysis were nominally significant in our study with the  
129 same directionality of effect. We have also compared our results to those described by  
130 HGI B2 analysis. We have observed significant associations in all previously described  
131 genome-wide loci but for the one on chr9, near the ABO locus (Supplementary File 3).

132 In addition to looking for replicated signals within our results, we ran two meta-  
133 analyses between BRACOVID hospitalization need and HGI B1 and B2 analysis. Adding  
134 our results to those provided by HGI B1 analysis disclosed a genome-wide association  
135 signal on chr6 that was only previously seen on the HGI B2 analysis, strengthening the  
136 confidence that this locus is indeed associated with Covid-19 severity (Supplementary  
137 File 4 and Supplementary Figure 4).

138 Our results meta-analyzing BRACOVID summary statistics with HGI B2 results  
139 supported most previously described genome-wide association loci. Namely, for Chr1  
140 near THBS3; Chr3 LZTFL1-CXCR6; Chr6 HLA-DRB1; Chr6 FOXP4; Chr11 MUC5B;  
141 Chr11 ELF5; Chr12 OAS2; Chr12 P2RX2; Chr17 CRHR1; Chr19 POLD1; Chr19  
142 TYK2; and Chr21 IFNAR2 regions we were able to observe concordant signals and no  
143 heterogeneity of effects. We want to emphasize the observation of significant  
144 heterogeneity for the previously described loci on Chr9 ABO locus; Chr21 SLC5A3; and  
145 Chr17 TAC4 suggesting that these may indeed be false-positive associations due to

146 population structure confounding, population-specific loci or loci associated with Covid-  
147 19 susceptibility and not severity (Supplementary File 5).

148

149 **Local association structure suggests a population specific effect for Chr1 DSTYK**  
150 **locus:**

151

152 The newly identified genome-wide significant locus near DSTYK was not  
153 observed as significantly associated in our HGI meta-analysis. By comparing the local  
154 association structure at the DSTYK locus, it is interesting to note that HGI B2 results also  
155 suggest a sub-genome-wide significant association at the same region (Figure 2A and B).  
156 However, albeit in the same recombination interval, there is no significant colocalization  
157 of the two genetic association structures, suggesting (1) the existence of more than one  
158 causal variants within this region or (2) that different LD structures among human  
159 populations may be confounding the association pattern at this locus.

160 To explore these scenarios, we have first analyzed the local association structure  
161 within the Brazilian sample stratifying by different degrees of European, African, and  
162 Native American genetic ancestry among participants. Curiously, although the signal is  
163 shared among the 3 groups, the strongest association signal was observed among the  
164 Brazilian individuals with high European genetic ancestry (Figure 2C). In addition to this,  
165 we have studied the pairwise LD between the most associated SNV in the Brazilian  
166 analysis (rs11240388), and the most associated SNV in HGI B2 analysis (rs9661015).  
167 Strikingly different world-wide pairwise LD patterns are observed among these 2  
168 markers (Figure 2F, <https://ldlink.nci.nih.gov/>). In our sample, the R-square between the  
169 two markers was 0.32 for participants with high European ancestry, and 0.11 for  
170 participants with high African ancestry.

171 Using only individuals with high European ancestry from our Brazilian dataset we  
172 delineated the associated haplotype boundaries between chr1:205131352 and  
173 chr1:205210329, defining a haplotype of approximately 79Kb. Looking into the  
174 haplotype structure of markers part of the associated haplotype we did not observe a  
175 significant difference in their pairwise LD structure among Brazilians with high  
176 European or high African ancestry. Using rs11240388 as a tag-SNP for our risk

177 haplotype, we defined world-wide allele frequencies for the risk haplotype (Figure 2G,  
178 <https://popgen.uchicago.edu/ggv/>). In our own sample, the frequency of the G allele (risk  
179 allele) was 48% for the entire sample, 42% in Brazilian individuals with high European  
180 ancestry and 59% in Brazilian individuals with high African ancestry. Despite having  
181 increased frequency in Brazilian individuals with high African ancestry, and as expected  
182 from the stratified local association plots, the haplotype was only significantly associated  
183 with Covid-19 severity in Brazilian individuals with high European genetic ancestry.  
184 Taken together these results suggest that the causal variant responsible for the association  
185 is present in a 79Kb haplotype derived from European populations.

186

187 **rs11240388 is also associated with several blood cell traits and with DSTYK and**  
188 **RBBP5 gene expression levels:**

189

190 Exploring the genomic annotations for the newly observed locus associated with  
191 Covid-19 hospitalization we used rs11240388 as a proxy of the risk haplotype for it is in  
192 complete LD with all the markers of the 79Kb haplotype. Rs11240388 is located at  
193 chr1:205208489 within the DSTYK gene region. The region appears to be delimited by a  
194 recombination region spanning the genes CNTN2, TMEM81, RBBP5, DSTYK, and  
195 TMCC2 and harbors several GeneHancer regulatory elements, most notably for CNTN2,  
196 DSTYK and TMCC2 (Figure 3A). Querying the MR-base database for other phenotypes  
197 previously shown to be associated with rs11240388, we observed 69 traits (Figure 3B).  
198 Of particular interest to Covid-19 biology are blood cell related traits and lung associated  
199 traits (Supplementary File 6).

200 Next, we explored the association structure between rs11240388 and GTEx gene  
201 expression levels in all tissues and for all genes within the region (Figure 3C). Results  
202 from GTEx suggest that rs11240388 is an eQTL in several tissues for CNTN2,  
203 TMEM81, RBBP5 and DSTYK. Colocalization analysis between GTEx association  
204 results and BRACOVID association results confirmed a significant colocalization for  
205 these genes in several different tissues, suggesting that the association between the 79Kb  
206 risk haplotype and Covid-19 severity might be mediated through gene expression  
207 modulation in the genes located in this genomic region (Figure 3D). Finally, we tested all

208 pair-wise possibilities for colocalization: BRACOVID signal, the 69 associated traits for  
209 the same region, and all GTEx gene-tissue results with significant colocalization (higher  
210 than 0.7) with BRACOVID (Supplementary Figure 5). This analysis provides several  
211 potential intermediate mediators that could explain the association between the identified  
212 risk haplotype and Covid-19 severity (Figure 3E).

213

## 214 **Discussion**

215 Covid-19 has been a world-wide pandemic for more than one year. The disease  
216 has killed more than 4 million individuals and has de-stabilized entire health systems as  
217 well as the world economy (2,9,10). One striking characteristic of the disease is the large  
218 inter-individual variability regarding its severity, ranging from asymptomatic to fatal.  
219 Although several identified factors influence disease severity, predicting the overall  
220 Covid-19 outcome in a particular individual remains a challenge (3–5).

221 Since the beginning of the pandemic, several groups worldwide have joined  
222 efforts to uncover host factors. The Host Genetic Initiative (HGI) included researchers  
223 and samples from different parts of the globe to focus on the inter-individual variability  
224 in disease presentation (8) and outcomes and the BRACOVID is currently the largest  
225 such effort in Latin America.

226 BRACOVID was designed to compare individuals with mild or asymptomatic  
227 Covid-19 with those with severe Covid-19 requiring hospitalization. This contrasts with  
228 most genetic studies on host determinants of Covid-19, that use individuals from the  
229 general population as their main control group. As such, BRACOVID has a more specific  
230 study protocol aimed at identifying host genetic determinants of disease severity, and not  
231 susceptibility.

232 The results here presented replicate most, but not all, previously reported genome-  
233 wide significant loci associated with Covid-19 severity. These came from 3 prior  
234 publications: The Severe Covid-19 GWAS Group, *GenOMICC* and HGI (6–8). It is  
235 interesting to note that despite the large sample sizes, none of these prior studies have in  
236 fact compared individuals with mild versus severe Covid-19. As such, our results confirm  
237 some of the initial results regarding Covid-19 genetic determinants in a more diverse  
238 population and, most importantly, within a study design to differentiate between markers



239 of susceptibility to infection and predictors of severity. Also of importance, although  
240 BRACOVID has participated in the meta-analysis of both *GenOMICC* and HGI, due to  
241 the analytical plan of these two prior publications, the contrast used by BRACOVID was  
242 not the one presented here. Specifically, in both *GenOMICC* and HGI, BRACOVID  
243 summary statistics were obtained by comparing hospitalized Covid-19 individuals with  
244 individuals from the general Brazilian population (the last group, not included in the  
245 present analysis) on what is called the HGI B2 analysis. Even in this latter situation,  
246 BRACOVID sample of hospitalized individuals used in the last version of HGI analysis,  
247 the one used in our meta-analysis with HGI B2, was of only 853 individuals, in contrast  
248 to the more than 3,500 hospitalized participants used in the present report.

249 Besides the importance of showing replications to previously identified loci in an  
250 admixed population, understanding loci that do not replicate might also be important in  
251 the follow-up efforts to identify population specific determinants of Covid-19 severity.  
252 One of the loci that was not replicated in our analysis is the association signal initially  
253 identified in the ABO locus. Our results, together with the verification of intense  
254 heterogeneity in signal described by HGI, suggest that the association initially described  
255 at this locus may be due to confounding by population structure or other higher order  
256 effects.

257 We have observed a new genome-wide significant locus associated with Covid-19  
258 in our population. It is comprised by a 79Kb haplotype located on chromosome 1 near the  
259 *DSTYK* gene. The association signal in our sample was only observed in individuals with  
260 high European ancestry and was not described by previous studies, suggesting it might  
261 harbor a population specific risk allele or haplotype. Indeed, leveraging data from the  
262 1000G project (11), we observed great inter-population differences in the haplotype  
263 frequency (although an apparently constant LD structure in the region among human  
264 populations).

265 Leveraging data from GTEx we were able to associate this haplotype with gene  
266 expression levels of at least 5 genes: *DSTYK*, *TMEM81*, *RBBP5*, *CNTN2* and *TMCC2*.  
267 In addition, exploring prior publicly available GWAS data we found significant  
268 associations between the identified risk haplotype and 69 different traits (8). From these,  
269 using colocalization analysis, we have observed significant colocalization between

270 BRACOVID results and several of these 69 traits, most strikingly a number of  
271 hematological traits, including immune cell morphology and number.

272 The main limitation of this study is the lack of a suitable replication sample for  
273 the GWAS results. The observed genome-wide significant locus near DSTYK needs to  
274 be replicated in an independent sample to continue to be considered a modulator of  
275 Covid-19 severity. The continuous effort to increase the sample size of the BRACOVID  
276 study may allow the identification genome-wide significant loci with decreased effect  
277 size.

278 Taken together, our analysis supports the hypothesis that the identified 79Kb  
279 haplotype on chromosome 1 may modulate Covid-19 relevant immunological traits  
280 through the gene expression of at least one of the genes with significant eQTLs in the  
281 region.

282

## 283 **Materials and methods**

284

### 285 **Study population**

286

287 The study sample belongs to the BRACOVID Study. For the present analysis we  
288 used the first 5,233 participants enrolled in the study.

289 The BRACOVID study was designed to identify and characterize genetic risk  
290 factors for Covid-19 severity in the Brazilian population. Participants were selected  
291 through two different ascertainment processes. First, Covid-19 patients were enrolled  
292 after hospitalization in one of the following tertiary care centers in the metropolitan area  
293 of Sao Paulo, Brazil: Instituto do Coração, and Instituto Central do Hospital das Clínicas  
294 da Faculdade de Medicina da Universidade de São Paulo. Non-hospitalized cases were  
295 selected by serological studies surveys for previous SARS-CoV-2 infection or SARS-  
296 CoV-2 PCR test among health professionals or the general population.

297 After signing an informed consent, a sample of whole-blood already collected for  
298 in-hospital biochemical analysis or SARS-CoV-2 serology was used for genomic DNA  
299 extraction.

300 Participants responded to a detailed questionnaire aimed at collecting  
301 demographic, risk factors, symptoms, and prior clinical information. For hospitalized  
302 participants we extracted the data from the hospital EHR.

303

304

### 305 **SNP Genotyping and Imputation**

306

307 Genomic DNA extraction has been previously described (12). BRACOVID DNA  
308 samples were genotyped using two different Axiom arrays: Axiom\_PMRA.r3 array  
309 (N=2605) or the Axiom\_sarscov array (N=2628) (ThermoFisher, Waltham, USA) and  
310 genotypes annotated using the array specific annotation file provided at the ThermoFisher  
311 website. Genotype calling was performed using Affymetrix Power Tools. Initial VCF file  
312 contained 850483 (for the PMRA array) and 779972 (for the Sars-Cov array) variants  
313 before quality control filtering.

314 Imputation was performed using the Haplotype Reference Consortium Michigan  
315 Imputation Server using the TOPMED reference haplotype panel as reference (for mixed  
316 samples). More specifically, the Michigan Imputation Server used Minimac4 to conduct  
317 imputation on 658,357 SNPs remaining after data quality control. After imputation data  
318 were exported in the standard PLINK format, downstream QC procedures and statistical  
319 analysis were conducted using the latest PLINK  
320 (<http://pngu.mgh.harvard.edu/~purcell/plink>) and R software packages (<http://www.r-project.org/>), installed on a Linux based computation resource. Specifically, imputation  
321 markers were kept if  $R^2 > 0.3$ , and minor allele frequency (MAF)  $> 0.01$ . A HWE p-  
322 value  $< 1 \times 10^{-20}$  was used to control for potential genotyping clustering problems.  
324 Genetic population structure was determined through PCA analysis after LD-pruning of  
325 associated markers (see also Statistical Analysis section). A total of 11,395,074 SNPs  
326 were used for genome-wide analyses, 10,981,197 for autosomal, and 413,877 for X-  
327 chromosomal analysis.

328

### 329 **GWAS Analysis**

330

331 We used a dichotomous category defined among SARS-CoV-2 participants. We  
332 have grouped individuals that did not require hospitalization into the control group  
333 (participants ascertained through population-based serological prevalence studies that did  
334 not have a prior history of hospitalization due to Covid-19) and hospitalized individuals  
335 (ascertained at the first hospital admission day) in the case group. Baseline categorical  
336 parameters are presented using frequencies (proportions), continuous parameters are  
337 presented using mean  $\pm$  SD.

338 Genome-wide association analyses were conducted using plink. We have  
339 conducted two analyses one without any further adjustment and one adjusting for the first  
340 four principal components, age, sex, and array type. The threshold for genome-wide  
341 significance was set to  $p < 5 \times 10^{-8}$ . Associations with  $p < 1 \times 10^{-6}$  were considered as  
342 suggestive and presented as a list of top associated SNPs.

343 Due to the high level of admixture and complex genetic population structure  
344 present in the Brazilian population we have conducted different sensitivity analysis  
345 taking into consideration the individual position on the PCA plot generated using the 2  
346 first principal components. Briefly, for the PCA-defined subgroup analysis we have k-  
347 means clustering with  $k=3$  and defined three different subgroups with higher European,  
348 African, and Native-American ancestries. Meta-analysis used a fixed-effect model and  
349 was calculated using plink –meta-analysis routine.

350 Local association plots were created using LocusZoom (13). Local linkage disequilibrium  
351 structure was determined using Haploview (14).

352

### 353 **PheWas analysis**

354

355 After identifying SNV that could be proxies of the associated haplotype we have  
356 used MR-base through the R package *ieugwasr* to search for prior GWAS studies  
357 reporting significant levels of association between a trait and the selected tag-SNV. The  
358 used p value cut-off level for this phewas analysis was  $p < 1 \times 10^{-5}$ .

359

### 360 **Colocalization analysis**

361

362 For colocalization analysis we have defined a window spanning 400Kb centered  
363 at the most associated variant in all regions classified as having a suggestive association  
364 signal. Information on all variants within this region was used for colocalization testing.  
365 We have used the R package *coloc* for colocalization analysis. Briefly, all genes residing  
366 in each selected region with their expression quantitative trait loci (eQTL) summary  
367 statistics available in GTEx V8.0 were sequentially tested for colocalization with the  
368 results obtained for Covid-19 severity association. As reference LD structure we used  
369 1000 genomes 2012 European LD matrix (our sample has approximately 80% European  
370 ancestry). Colocalization was tested against all 48 tissues available in GTEx V8.0 and  
371 against all traits with significant association to the selected variant (as described in the  
372 phewas section). We used a threshold of H4 (the posterior probability that a single causal  
373 variant, or haplotype, could explain the local association pattern of both tested traits) >  
374 0.7 as evidence for significant colocalization.

375

### 376 **Acknowledgements**

377 The BRACOVID project received an unrestricted financial support from JBS S.A.  
378 Company, Sao Paulo, Brazil and from the general public under the HC-COMVIDA  
379 crowdfunding scheme (<https://viralcure.org/c/hc>). The funds were managed by Fundação  
380 Zerbini and Fundação Faculdade de Medicina, respectively. We are thankful for the infra-  
381 structure support for sample collection provided by the HCFMUSP Covid-19 Study  
382 Group. We are also grateful to the Host Genetic Initiative for making their data publicly  
383 available (full acknowledgements can be found at  
384 <https://www.covid19hg.org/acknowledgements/>).

385

### 386 **Conflict of interest**

387 The authors certify that they have NO affiliations with or involvement in any  
388 organization or entity with any financial interest in the subject matter or materials  
389 discussed in this manuscript.

390

## 391 **References**

- 392 1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from  
393 Patients with Pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727–33.
- 394 2. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE)  
395 at Johns Hopkins University (JHU) [Internet]. 2021. Available from:  
396 <https://coronavirus.jhu.edu/map.html>
- 397 3. Morens DM, Fauci AS. Emerging Pandemic Diseases: How We Got to COVID-  
398 19. *Cell.* 2020;183(3):837.
- 399 4. Casanova J-L, Abel L. Lethal Infectious Diseases as Inborn Errors of Immunity:  
400 Toward a Synthesis of the Germ and Genetic Theories. *Annu Rev Pathol.* 2021  
401 Jan;16:23–50.
- 402 5. Casanova JL, Abel L. The human genetic determinism of life-threatening  
403 infectious diseases: genetic heterogeneity and physiological homogeneity? *Hum*  
404 *Genet* [Internet]. 2020;139(6–7):681–94. Available from:  
405 <https://doi.org/10.1007/s00439-020-02184-w>
- 406 6. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N*  
407 *Engl J Med.* 2020;383(16):1522–34.
- 408 7. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al.  
409 Genetic mechanisms of critical illness in COVID-19. *Nature.* 2021;591(7848):92–  
410 8.
- 411 8. Ganna A, Unit TG, General M. The COVID-19 Host Genetics Initiative, a global  
412 initiative to elucidate the role of host genetic factors in susceptibility and severity  
413 of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet.* 2020;28(6):715–8.
- 414 9. Weinstein M, Skinner J. Covid-19 — Implications for the Health Care System  
415 David. *N Engl J Med.* 2011;362(5):567–71.
- 416 10. Allain-Dupré D, Chatry I, Michalun V, Moisisio A. The territorial impact of

- 417 COVID-19□: managing the crisis across levels of government. OECD Tackling  
418 Coronavirus. 2020. p. 2–44.
- 419 11. Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, et  
420 al. A map of human genome variation from population-scale sequencing. *Nature*.  
421 2010;467(7319):1061–73.
- 422 12. Preparation S, No DP. Axiom™ 2.0 gDNA Sample Preparation [Internet]. p. 2–5.  
423 Available from: [https://www.thermofisher.com/us/en/home/technical-](https://www.thermofisher.com/us/en/home/technical-resources/contact-us.html?supportType=TS)  
424 [resources/contact-us.html?supportType=TS](https://www.thermofisher.com/us/en/home/technical-resources/contact-us.html?supportType=TS)
- 425 13. Boughton AP, Welch RP, Flickinger M, VandeHaar P, Taliun D, Abecasis GR, et  
426 al. LocusZoom.js: Interactive and embeddable visualization of genetic association  
427 study results. *Bioinformatics*. 2021 Mar;
- 428 14. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD  
429 and haplotype maps. *Bioinformatics*. 2005 Jan;21(2):263–5.

430

### 431 **Supporting Information**

#### 432 **Tables:**

433 **Table 1.** Demographic, clinical and laboratory characteristics of studied subjects  
434 according to hospitalization need.

Patients' characteristics	Total n (%)	Hospitalization n (%)	Non-hospitalization n (%)
	5233 (100)	3533 (67.5)	1700 (32.5)
<i>Demographic characteristics</i>			
<i>Age categories</i>			
<18	5 (0.1)	5 (0.1)	0 (0)
18–49 (average: 36)	2308 (45)	986 (28)	1322 (80.7)
50–64 (average: 57)	1423 (27.5)	1156 (32.8)	267 (16.3)
>65 (average: 74)	1424 (27.6)	1371 (39)	49 (3)
<i>Sex</i>			
Female	2838 (54)	1626 (46)	1212 (71.4)
Male	2391 (45.8)	1906 (54)	485 (28.6)
<i>Race/ethnicity</i>			
White	3409 (66)	2453 (70.8)	956 (56.2)
Brown	1183 (23)	684 (19.7)	499 (29.4)
Black	445 (8.6)	244 (7)	201 (11.8)
Native Brazilian	5 (0.1)	1 (0.02)	4 (0.2)
Asian	69 (1.3)	39 (1.1)	30 (1.8)
Other	6 (0.1)	4 (0.1)	2 (0.1)
<i>Health characteristics</i>			
Hypertension	2185 (42)	1937 (55.3)	248 (14.7)
Cardiovascular Disease	1021 (20)	996 (29.6)	25 (1.5)
ACEi users	431 (9.5)	395 (13.7)	36 (2.2)
Asthma	239 (4.6)	172 (4.9)	67 (4)
Tabagism	1002 (19.4)	833 (24)	169 (10)
COPD	219 (4.2)	213 (6.1)	6 (0.4)
Diabetes melitus	1329 (25.7)	1238 (35.4)	91 (5.4)
HIV	35 (0.7)	33 (0.9)	2 (0.1)
CKD	449 (8.7)	441 (12.6)	8 (0.5)
<i>BMI statusd</i>			
Obese (BMI of 30 kg/m <sup>2</sup> or greater)	1186 (32.2)	735 (36.4)	447 (26.8)
Overweight (BMI of 25–29.9 kg/m <sup>2</sup> )	1341 (36.4)	683 (33.8)	659 (39.6)
Normal (BMI of 20–24.9 kg/m <sup>2</sup> )	984 (26.7)	489 (24.2)	496 (29.8)
BMI of 20 kg/m <sup>2</sup> or lower	176 (4.8)	113 (5.6)	63 (3.8)
Cancer	191 (3.7)	170 (4.9)	21 (1.2)
Autoimmune diseases	132 (2.6)	101 (2.9)	31 (1.8)
<i>Case severity</i>			
Hospitalization	3533 (67.5)	3533 (100)	0 (0)
ICU	2186 (41.8)	2186 (62.6)	0 (0)
Oxygen	2621 (50.1)	2619 (75.1)	2 (0.1)
No-invasive ventilation	1263 (24.1)	1263 (36.2)	0 (0)
Mechanical ventilation	1521 (29.1)	1521 (43.6)	0 (0)
Dialysis	511 (9.8)	511 (14.8)	0 (0)
<i>Vaccination status</i>			
Yes (2 doses)	856 (16.4)	60 (1.7)	796 (46.8)
Yes (1 dose)	102 (1.9)	58 (1.6)	44 (2.6)
No	4275 (81.7)	3415 (96.7)	860 (50.6)

435  
436  
437  
438  
439

**ACEi: Angiotensin-converting enzyme inhibitors; BMI: Body Mass Index; CKD: Chronic kidney disease; COPD: Chronic obstructive pulmonary disease; HIV: Human immunodeficiency virus; ICU: Intensive care unit; NA: Not available.**

440 **Table 2.** Genome-wide significant and suggestive loci identified in BRACOVID Covid-  
441 19 severity analysis.



CHR	BP	SNP	Effect Allele	P	OR	Genes	
1	205208489	chr1:205208489:G:A	G	3.99E-08	1.3516	RBBP5, DSTYK, TMCC2	
2	45584979	chr2:45584979:T:C	C	4.06E-07	0.7509	SRBD1	
4	25613313	chr4:25613313:G:A	A	3.21E-07	1.4647	SLC34A2	
5	21768177	chr5:21768177:G:A	A	5.02E-07	1.3263	GUSBP1, CDH12	
10	22439346	chr10:22439346:T:C	C	5.96E-07	0.549	PIP4K2A	

442

443 **Figures:**

444 **Figure 1.** Manhattan plot of BRACOVID results for Covid-19 hospitalization. Results  
445 from a logistic regression model adjusted for the first 4 PCs, age, and sex for each used  
446 genotyping array were meta-analyzed using a fixed-effect meta-analysis.

447

448 **Figure 2.** Association and LD structure at Chr1 DSTYK associated locus. A. Local  
449 association structure of BRACOVID hospitalization analysis; B. Local association  
450 structure of HGI B2 analysis; C. Local association structure for BRACOVID European  
451 cluster; D. Local association structure for BRACOVID African cluster; E. Local  
452 association structure for BRACOVID Native American cluster; F. World-wide  
453 distribution of pairwise LD between rs11240388 (local maximum for BRACOVID  
454 analysis) and rs9661015 (local maximum for HGI B2 analysis); G. 1000G allele  
455 frequency distribution for rs11240388.

456 **Figure 3.** Annotation and functional characterization of Chr1 DSTYK associated region.  
457 A. Genomic anatomy of associated region. Blue highlight – 79Kb associated haplotype.  
458 B. Phewas results and colocalization analysis with BRACOVID summary statistics.  
459 Colocalization values are depicted as a function of their H3 and H4 posterior  
460 probabilities. Positive values (blue spectrum) for higher H4 values than H3, negative  
461 values (red spectrum) for higher H3 than H4. Inner circle represents study/database of  
462 trait being colocalized with BRACOVID results. Black: Datasets that satisfy minimum  
463 requirements imported from the EBI database of complete GWAS summary data; red:

464 eqtlGen 2019 results; green: GWAS summary datasets generated by many different  
465 consortia that have been manually collected and curated, initially developed for MR-Base  
466 (round 2); dark blue: Complete GWAS summary data on protein levels as described by  
467 Sun et al 2018; magenta: Complete GWAS summary data on brain region volumes as  
468 described by Elliott et al 2018; pink: Neale lab analysis of UK Biobank phenotypes,  
469 round 2; C. Association structure between rs11240388 and genes within its genomic  
470 region using 42 tissues in GTEx V8. Red rectangles are for association p values below  
471 0.01. Numbers within cells represent the normalized mean expression value within tissue  
472 for gene; D. Colocalization results between BRACOVID summary statistics for  
473 rs11240388 region and different GTEx V8 gene/tissue combinations. Shown are only the  
474 most significant H4 probability results; E. Causal diagram derived from pairwise  
475 comparisons with significant colocalization at rs11240388 region.

476

477 **Supplementary Figures:**

478 **Supplementary Figure 1.** Local association plots for the 4 suggestive associated loci.  
479 Suggestive association was defined as a p-value between  $1 \times 10^{-6}$  and  $5 \times 10^{-8}$ . Genome  
480 positions in Hg19.

481 **Supplementary Figure 2.** K-means clustering ( $k=3$ ) of BRACOVID participants using  
482 the first and second PCs. Orange triangle - European cluster; Green circle - African  
483 cluster; Purple square - Native American cluster.

484 **Supplementary Figure 3.** Manhattan plots of ancestry clusters and trans-ethnic meta-  
485 analysis. A. European cluster; B. African cluster; C. Native American cluster; D. Native  
486 American cluster.

487 **Supplementary Figure 4.** Local association structure for Chr6 FOXP4 region.

488 **Supplementary Figure 5.** Pairwise colocalization results for all pairwise combinations  
489 of significant GTEx gene/tissue pairs, traits associated with rs11240388 in the phewas  
490 analysis and BRACOVID hospitalization results for the rs11240388 region. Red

491 spectrum for colocalization results where H3 posterior probability > H4 posterior  
492 probability. Blue spectrum for colocalization results where H4 posterior probability > H3  
493 posterior probability.

494

495 **Supplementary files**

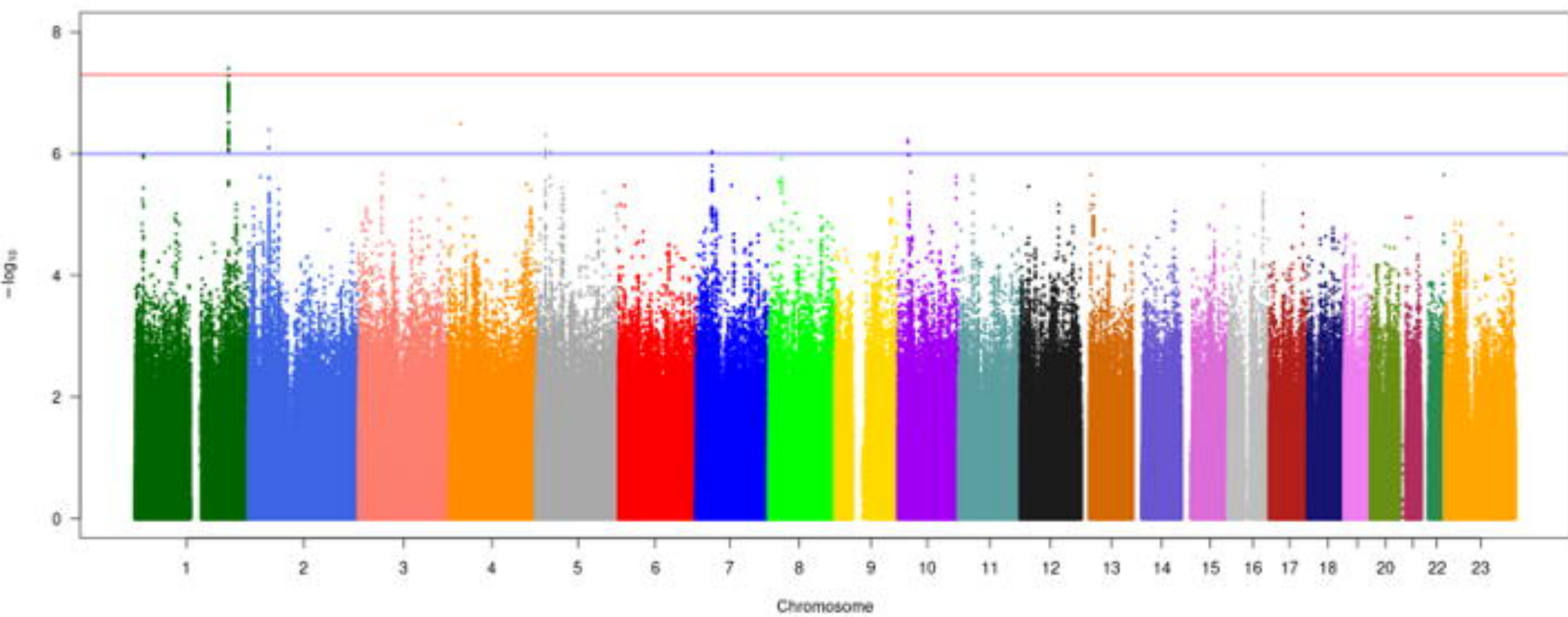
496 **Supplementary File 1.** Replication results from BRACOVID for HGI B1 analysis  
497 genome-wide significant loci.

498 **Supplementary File 2.** Replication results from BRACOVID for HGI B2 analysis  
499 genome-wide significant loci.

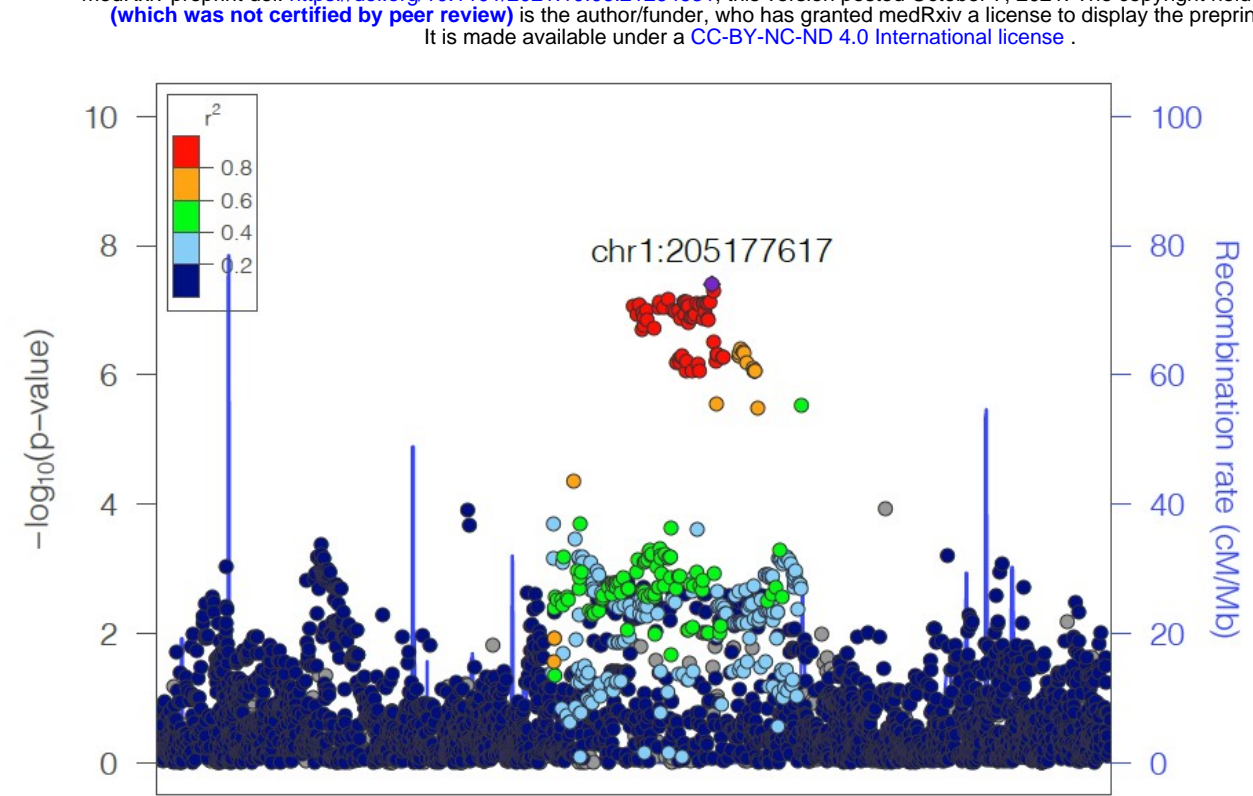
500 **Supplementary File 3.** Genome-wide significant hits from meta-analysis BRACOVID  
501 versus HGI B1 model.

502 **Supplementary File 4.** Genome-wide significant hits from meta-analysis BRACOVID  
503 versus HGI B2 model.

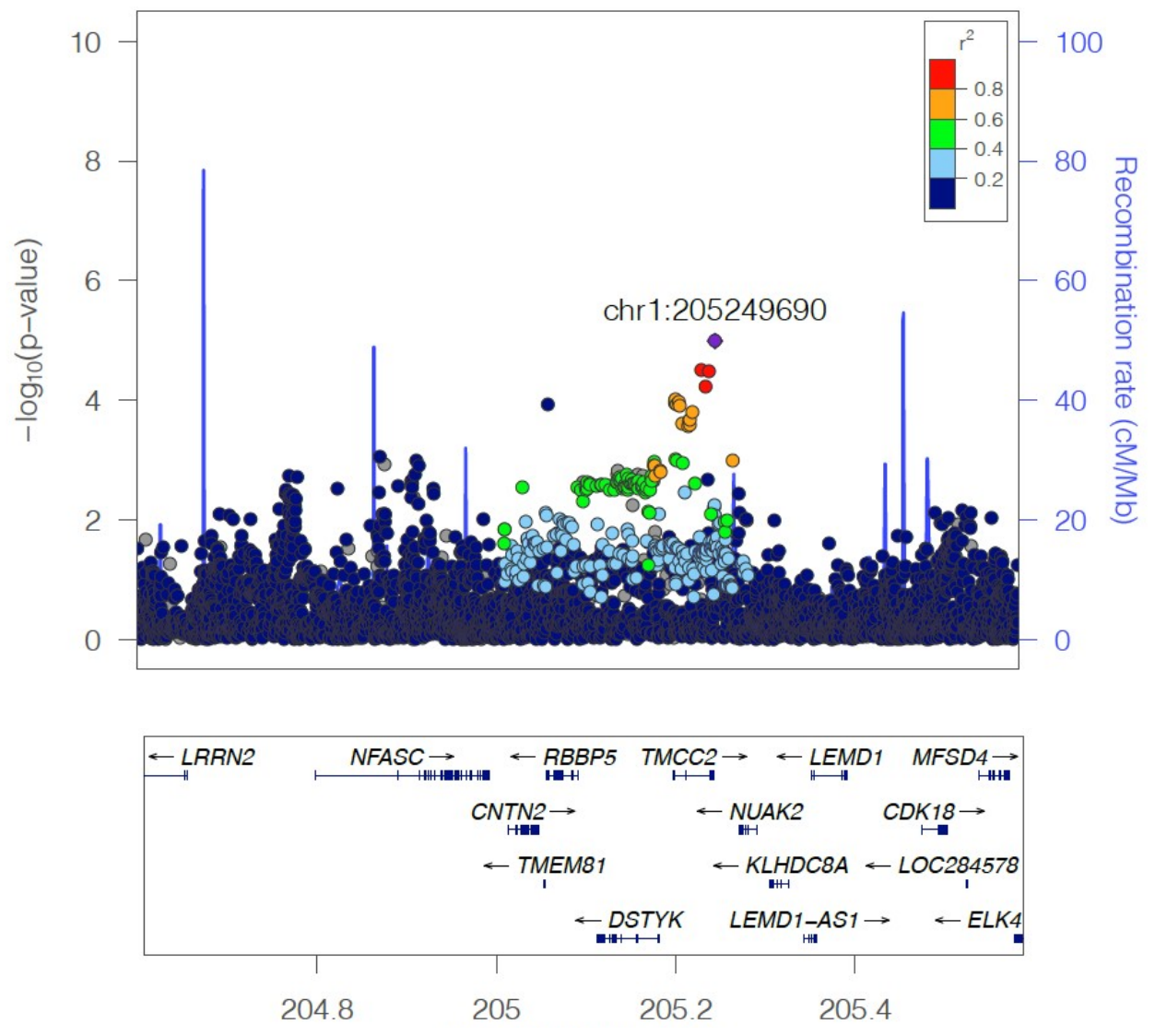
504



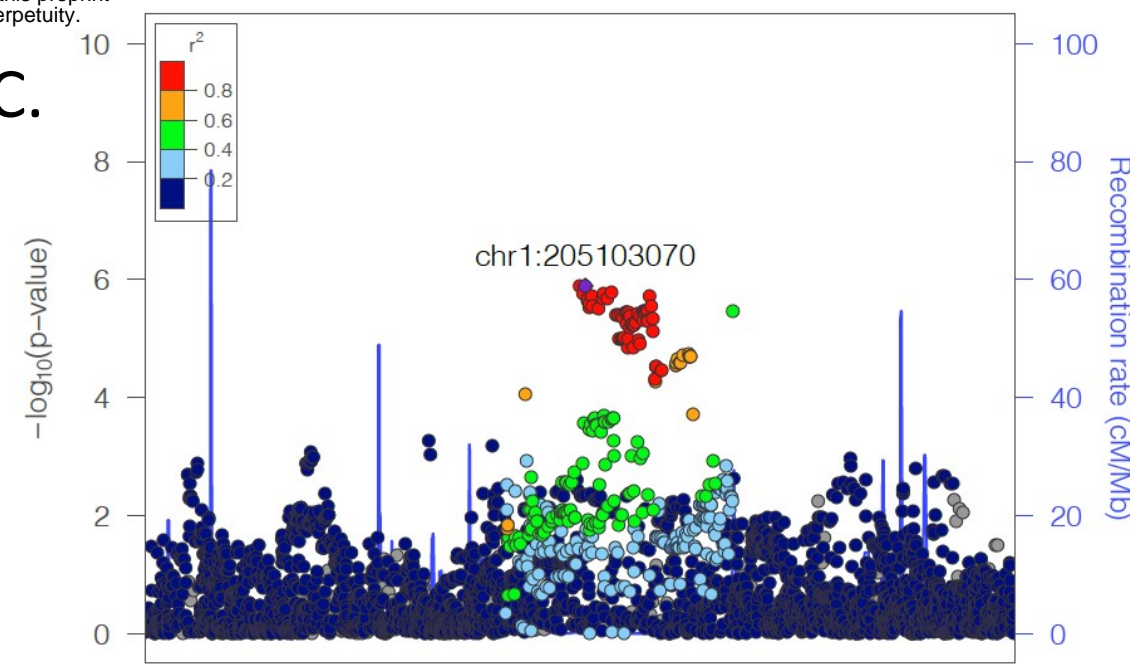
A.



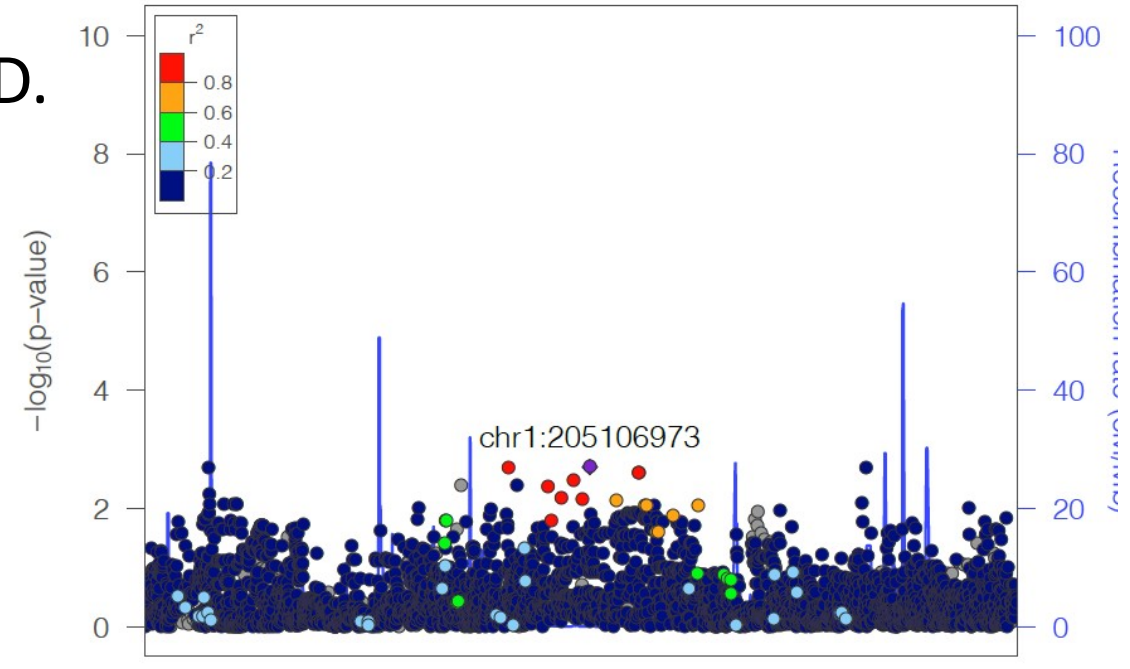
B.



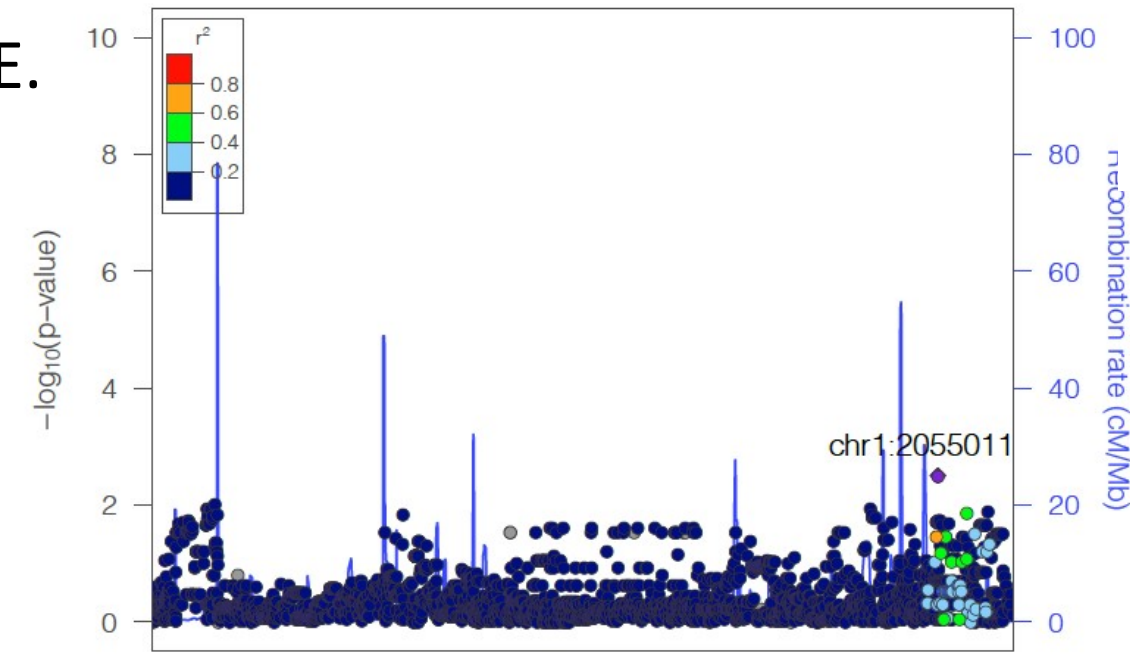
C.



D.



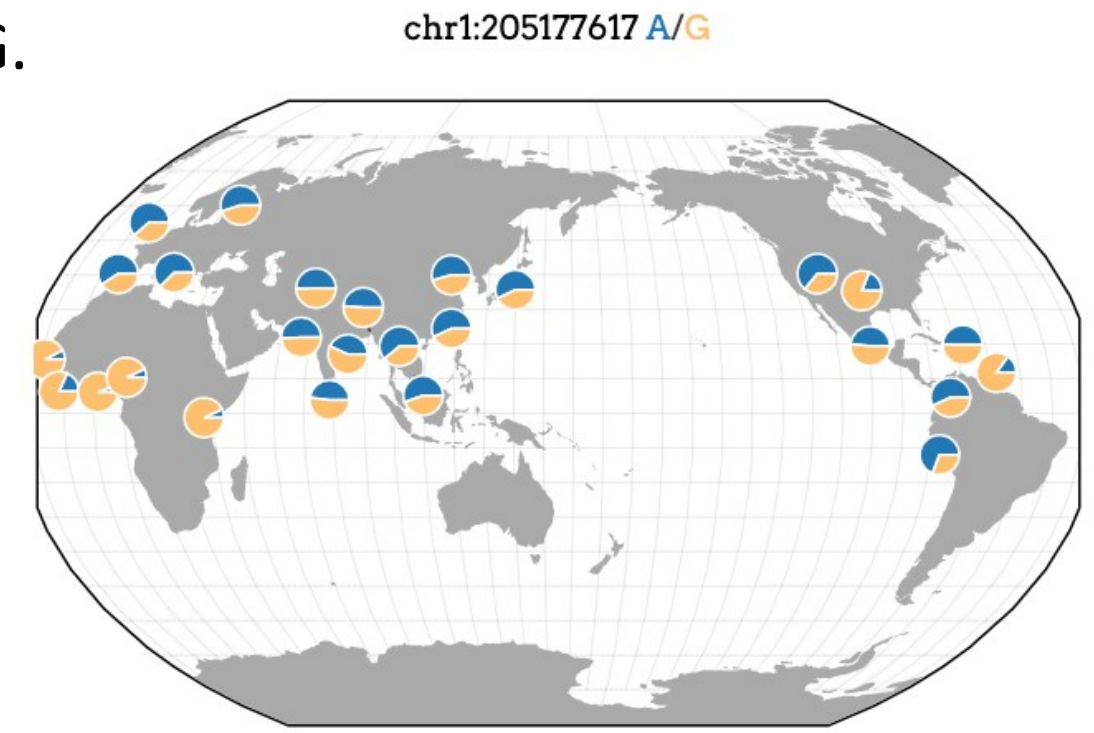
E.

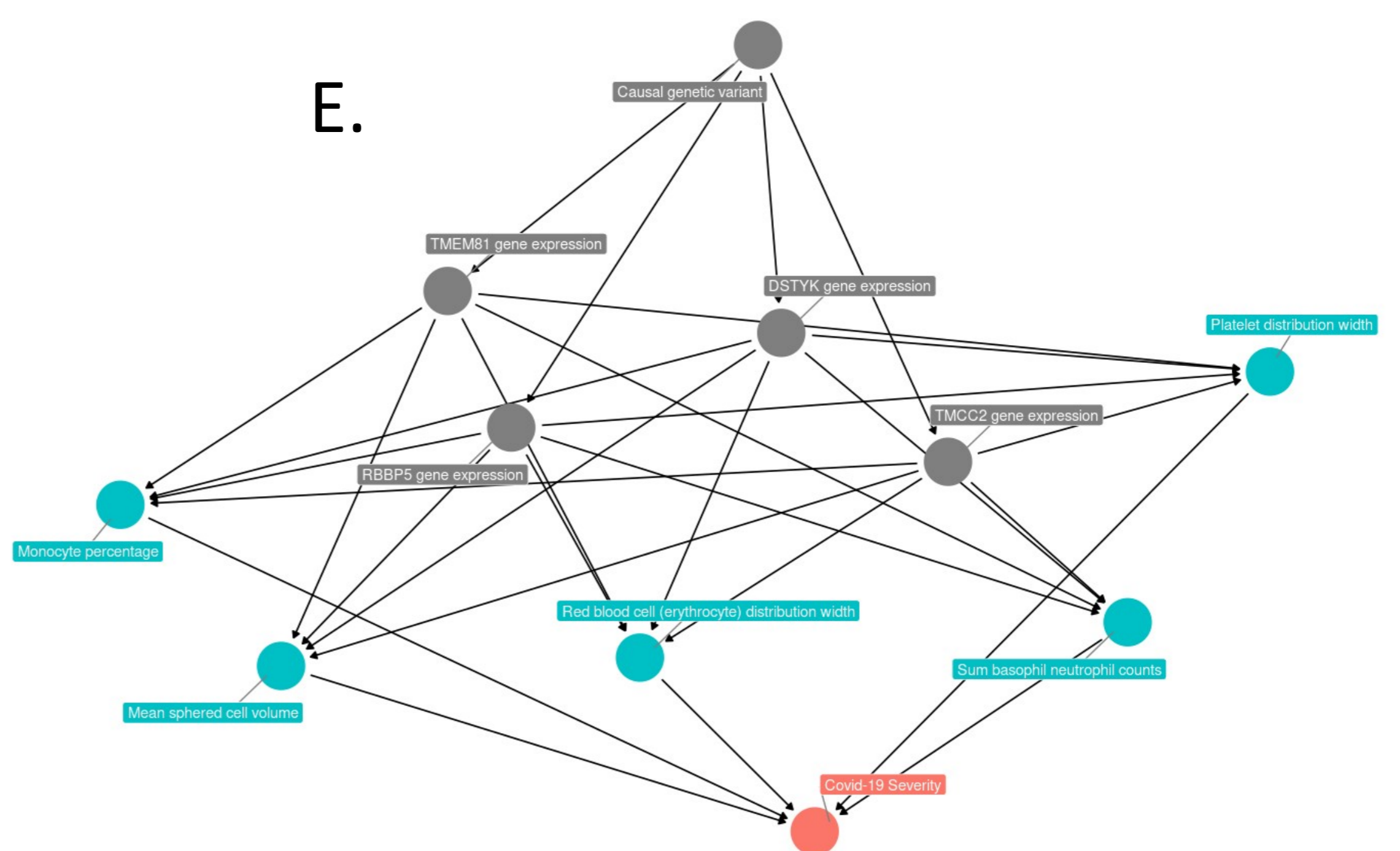
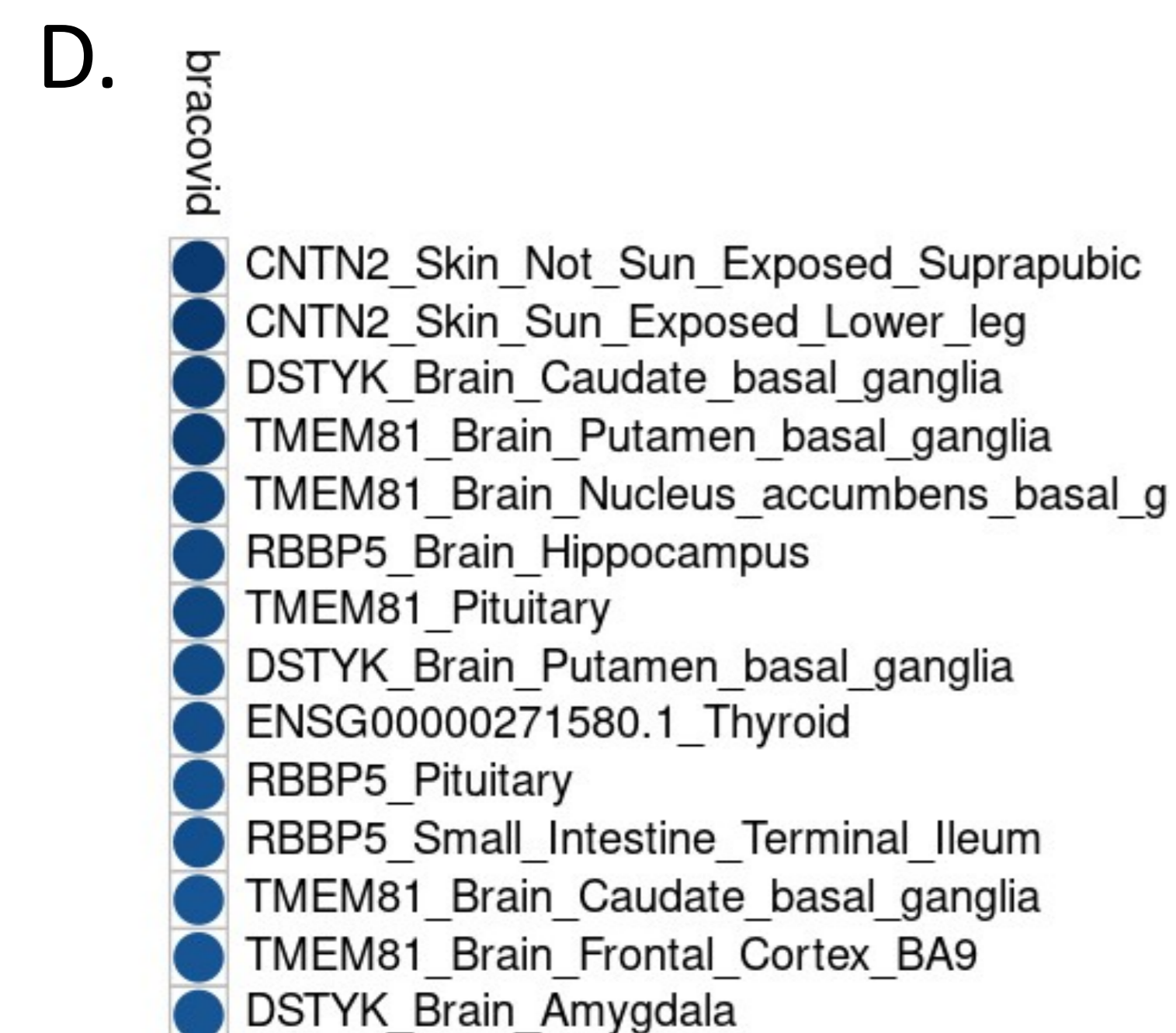
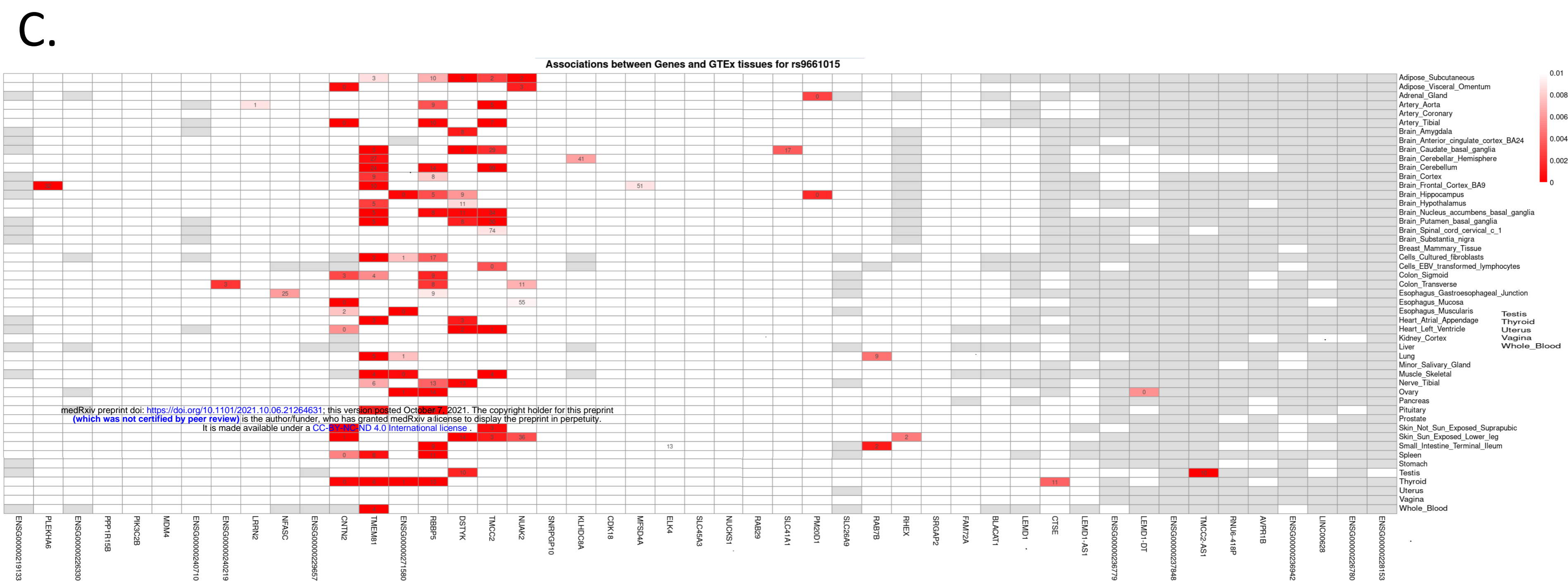
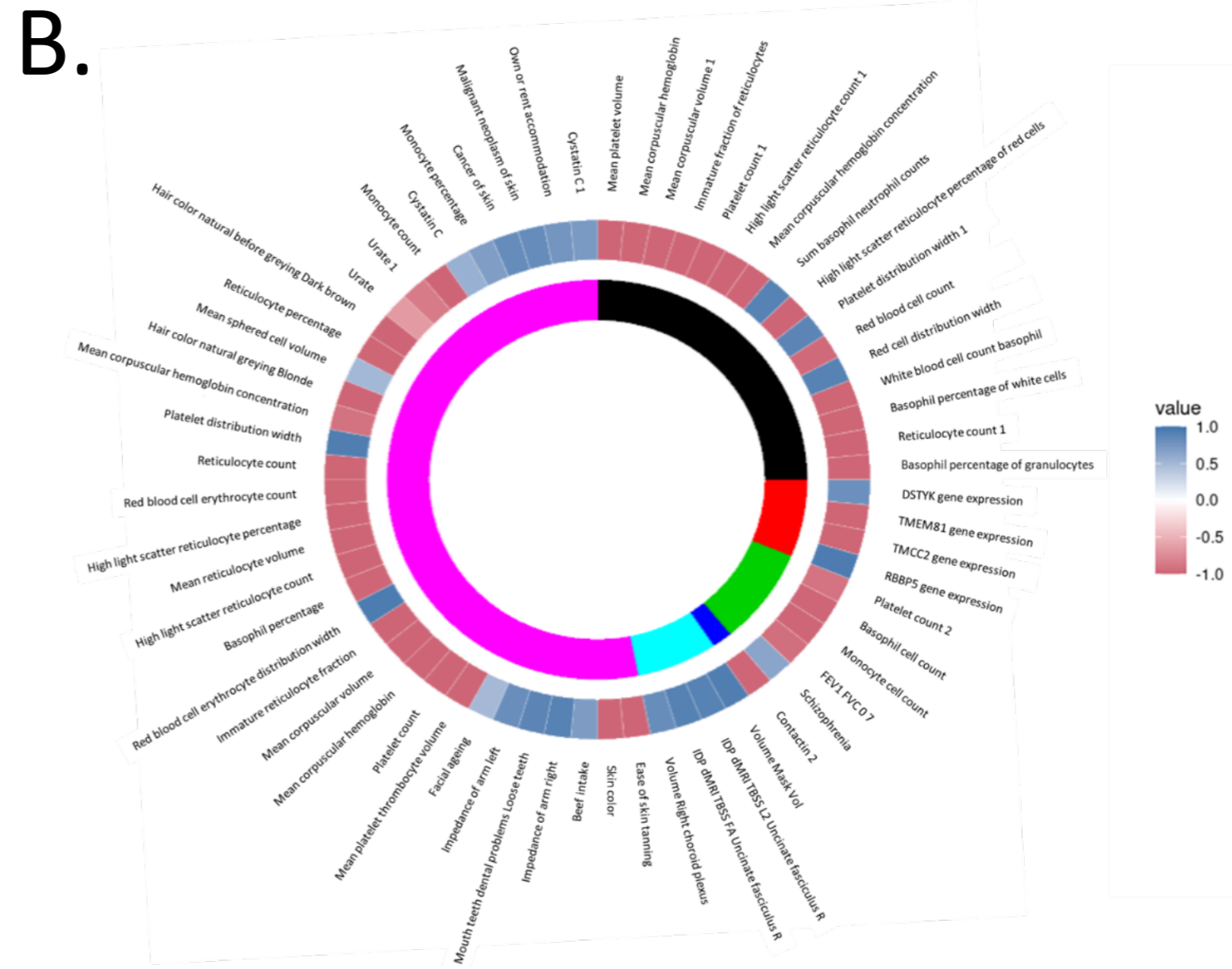
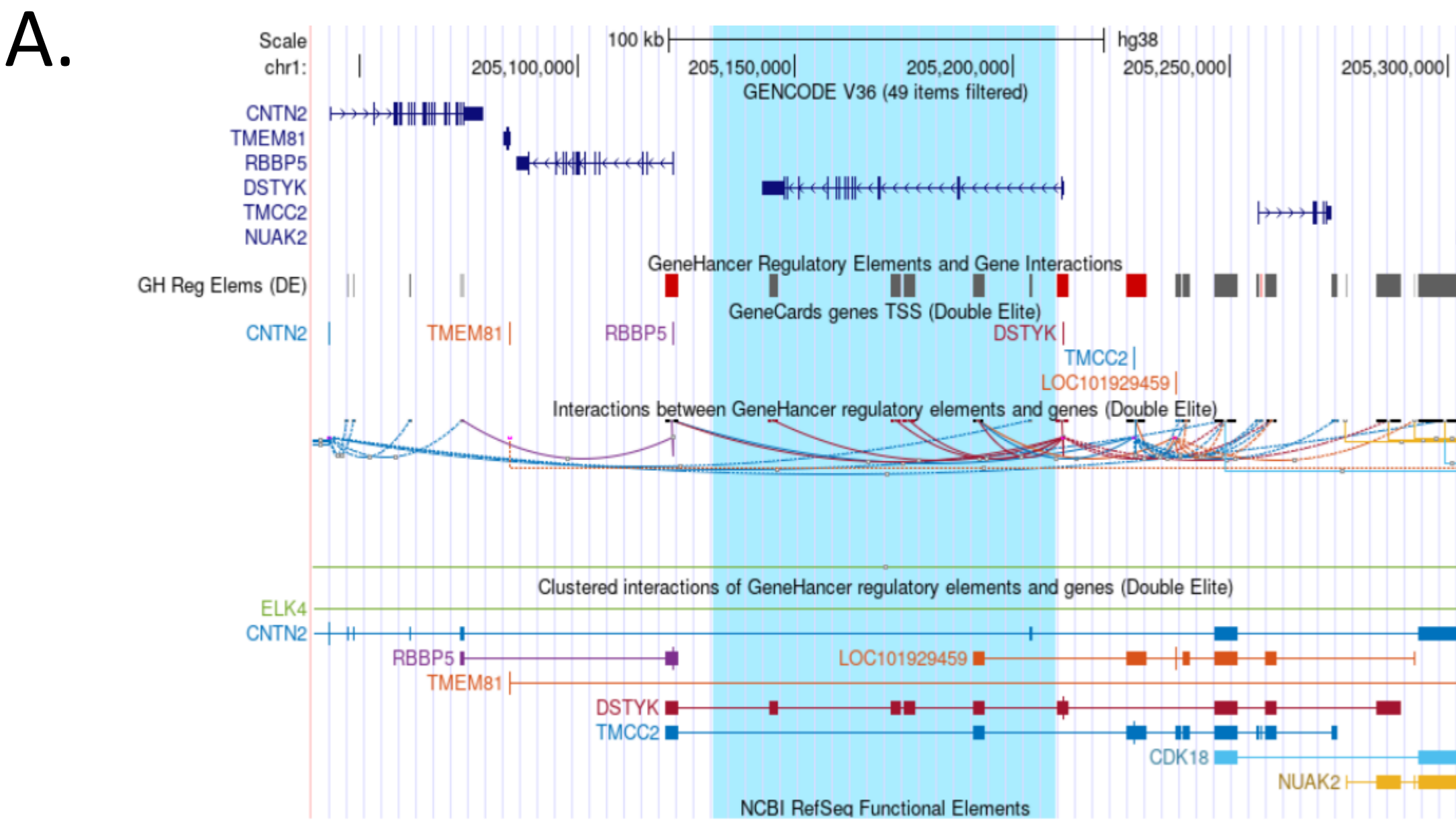


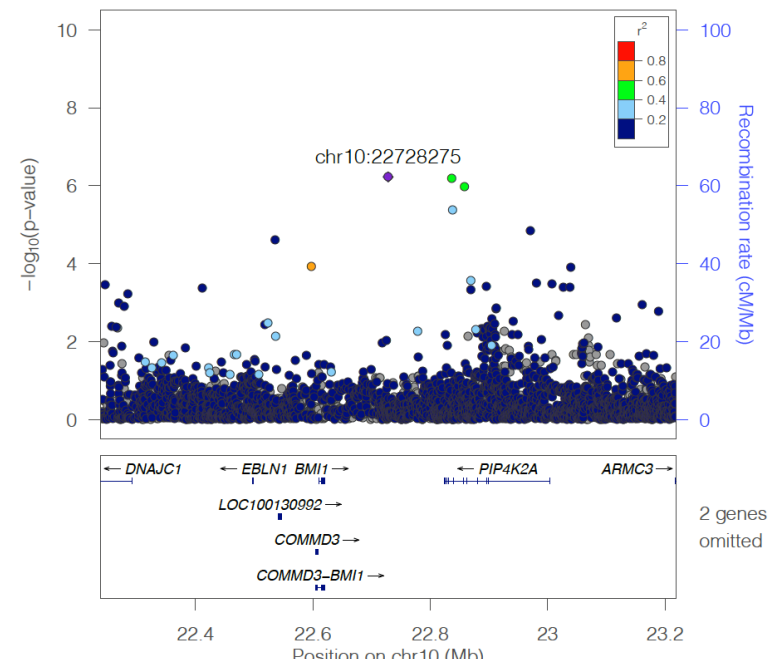
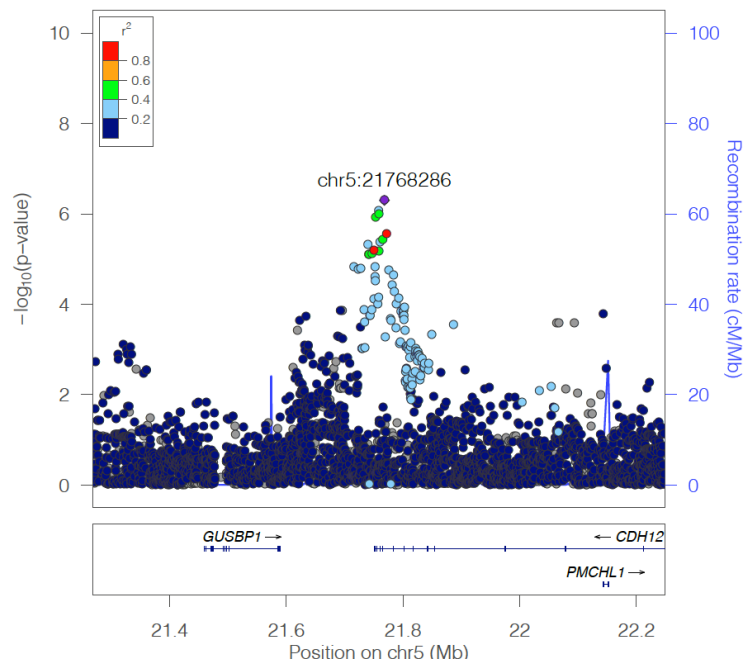
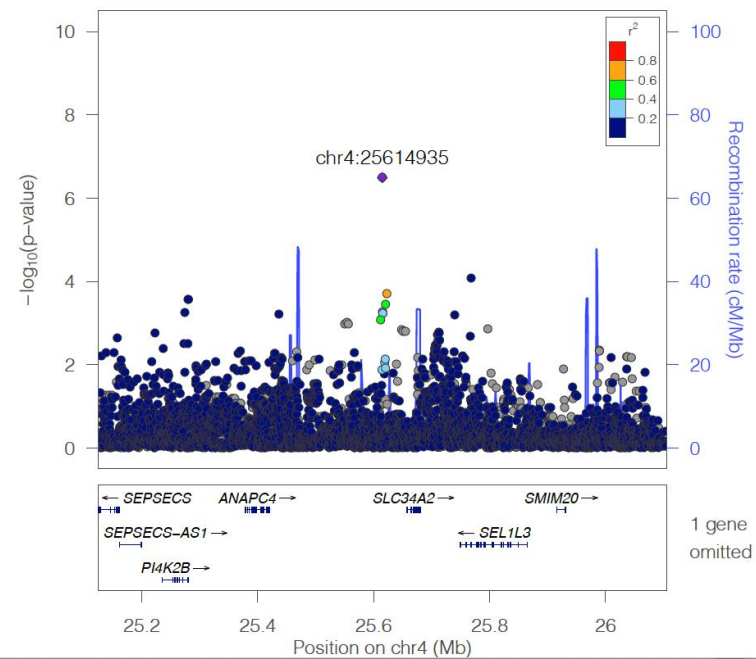
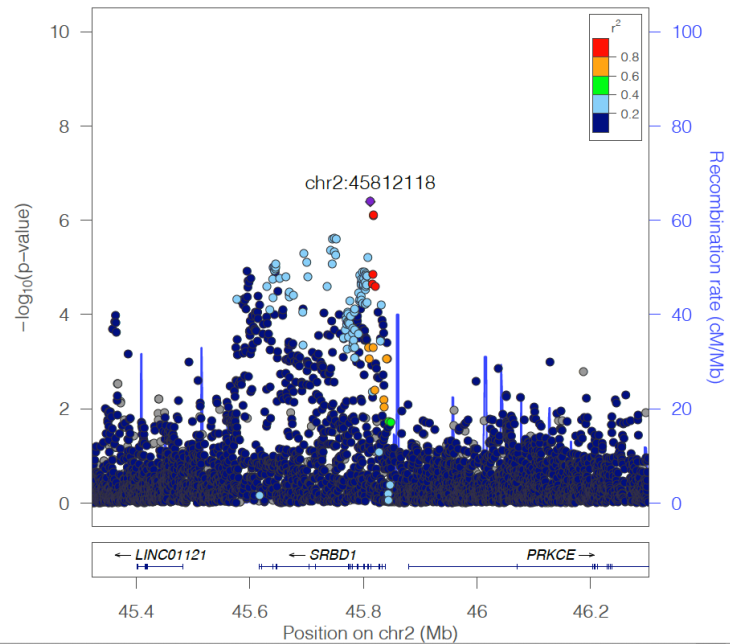
F.



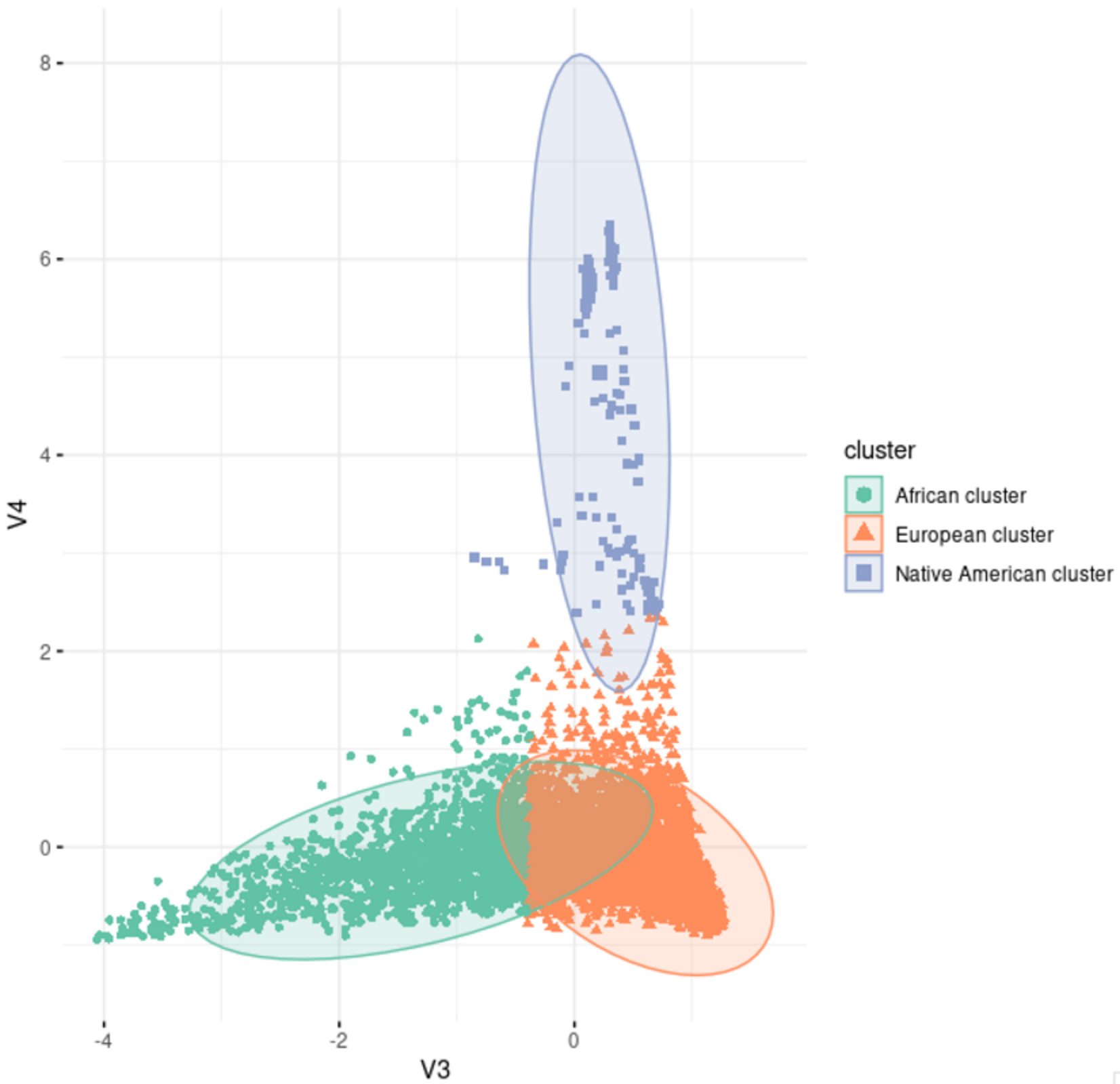
G.



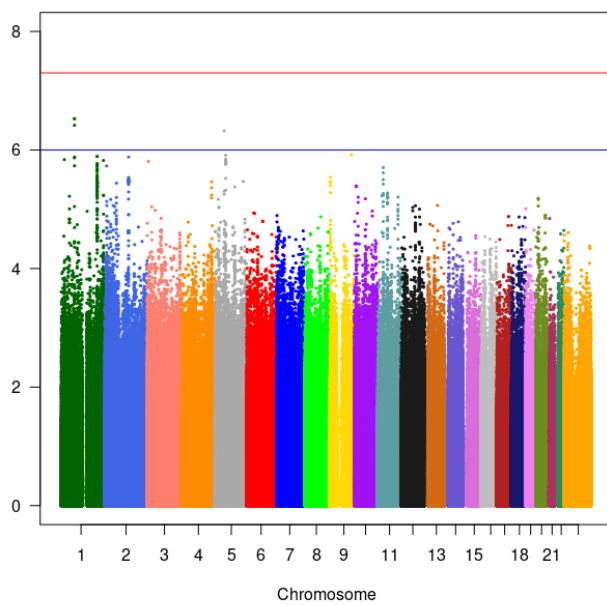
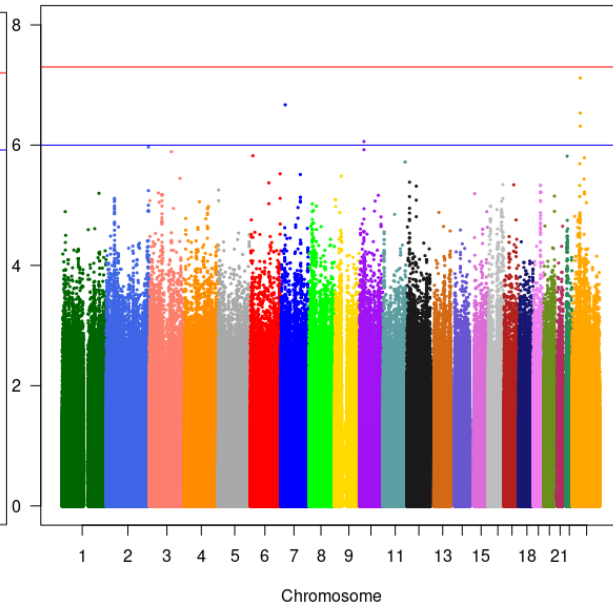
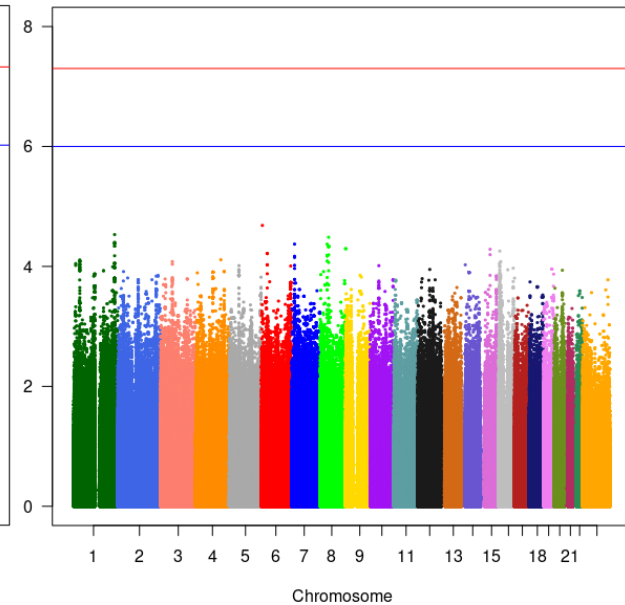
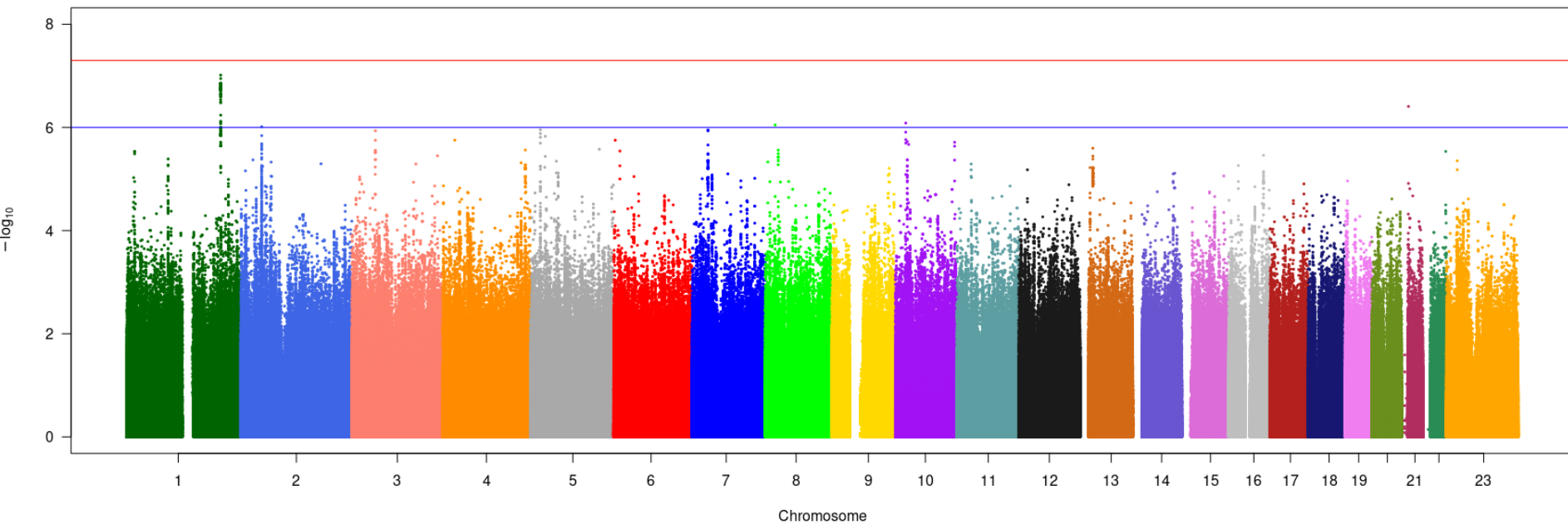




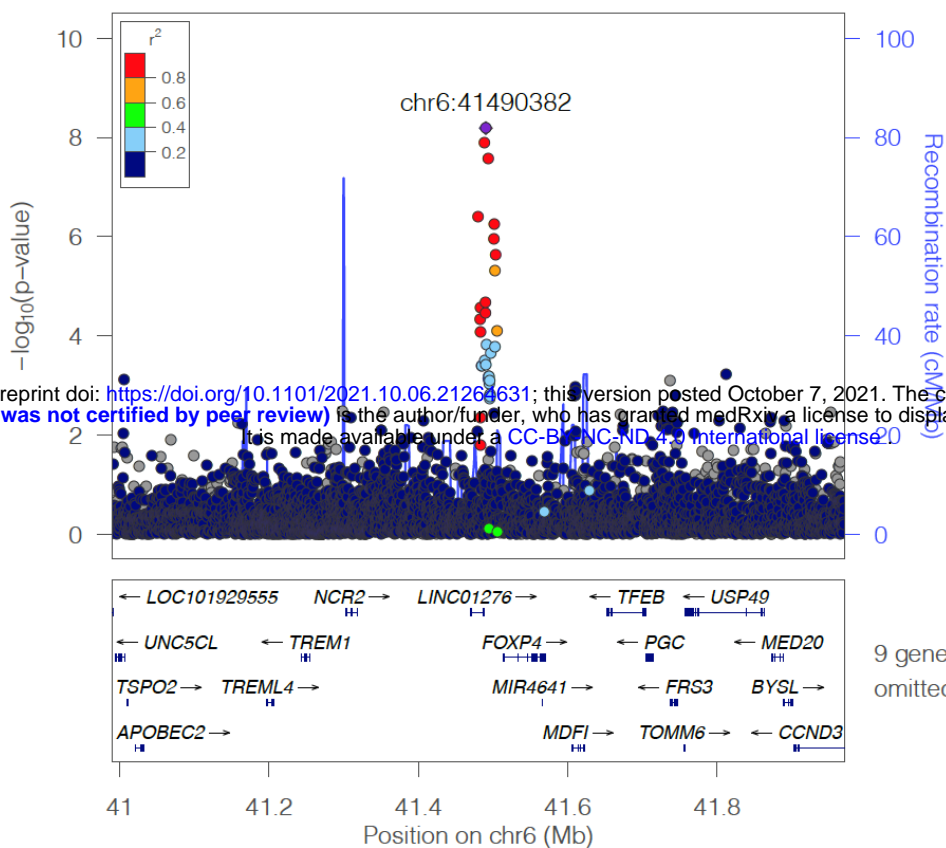
Cluster plot



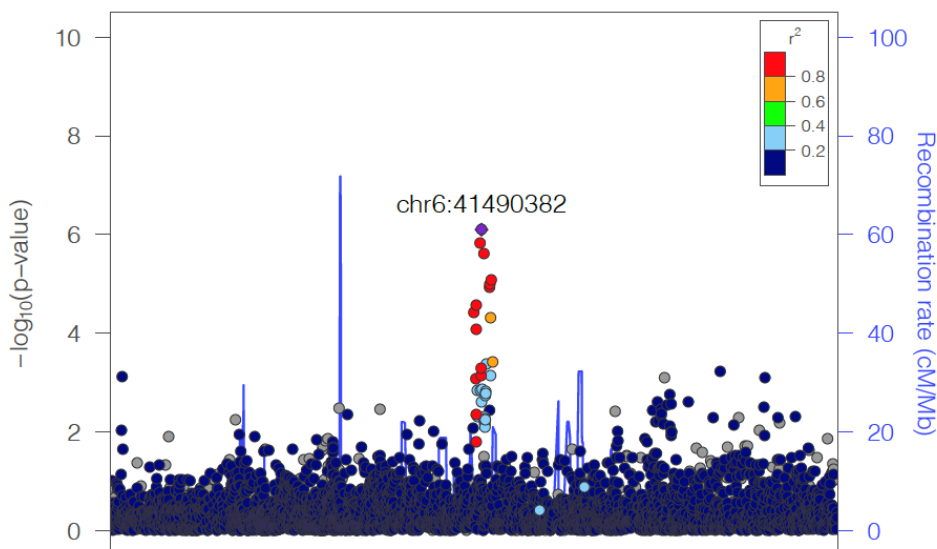


**A.****B.****C.****D.**

## Meta-analysis



## HGI B1 model



## BRACOVID Hospitalization

